4Д

УДК 622.276.5.001.42 + 519.23/.25

С.В. Денисов [1], ***К.А. Сидельников*** [2]
[1]Уфимский государственный нефтяной технический университет,
г. Уфа, Россия
[2]Ижевский государственный технический университет, г. Ижевск, Россия
sk-albion@yandex.ru

Верификация модели пласта-коллектора по данным гидродинамических исследований скважин с использованием доверительных интервалов

Представлена базовая процедура нелинейного регрессионного анализа. Использован байесовский вывод, при котором вся информация о параметрах пласта выражается в терминах вероятности, а неопределенность, связанная с оценками этих параметров, может определяться количественным образом. Исследованы проблемы, возникающие при использовании доверительных интервалов в рамках байесовского вывода для дискриминантного анализа различных моделей-кандидатов.

Алгоритм нелинейной регрессии

Цель нелинейного регрессионного анализа на основе метода наименьших квадратов (МНК) состоит в минимизации суммы квадратов разностей между измеренными и вычисленными значениями давления в рамках выбранной модели пласта:

$$E = \sum_{i=1}^{n} \left(y_i - F\left(\mathbf{и},\, x_i\right)\right)^2, \qquad (1)$$

где E – целевая функция; F – функция, описывающая модель пласта; $\boldsymbol{\theta}$ – неизвестные параметры пласта; x_i – зависимая переменная (время); y_i – независимая переменная (давление); n – число данных.

Функция F, а значит, и целевая функция E – это обычно нелинейная функция неизвестных параметров пласта. Вследствие этого нужно многократно обновлять неизвестные параметры до тех пор, пока целевую функцию нельзя будет сделать еще меньше. Метод Гаусса-Маквардта со штрафными функциями и специальной техникой интерполяции и экстраполяции представляет собой модификацию метода Ньютона. Математической базой метода Ньютона является теорема Тейлора. В методе Ньютона целевая функция аппроксимируется квадратичной моделью путем отбрасывания членов ряда Тейлора вблизи начального приближения набора неизвестных параметров:

$$E^* = E\Big|_{\boldsymbol{\theta}^0} + \sum_{j=1}^{m}\left(\theta_j - \theta_j^0\right)\frac{\partial E}{\partial \theta_j}\bigg|_{\boldsymbol{\theta}^0} + \sum_{j=1}^{m}\sum_{k=1}^{m}\left(\theta_j - \theta_j^0\right)\left(\theta_k - \theta_k^0\right)\frac{\partial^2 E}{\partial \theta_j \partial \theta_k}\bigg|_{\boldsymbol{\theta}^0}, \qquad (2)$$

где

$$\frac{\partial E}{\partial \theta_j}\bigg|_{\boldsymbol{\theta}^0} = -2\sum_{i=1}^{n}\left(y_i - F\left(\theta^0,\, x_i\right)\right)\frac{\partial F}{\partial \theta_j}\bigg|_{\boldsymbol{\theta}^0}; \qquad (3)$$

$$\left.\frac{\partial^2 E}{\partial\theta_j \partial\theta_k}\right|_{\boldsymbol{\theta}^0} = -2\sum_{i=1}^{n}\left(-\left.\frac{\partial F}{\partial\theta_j}\frac{\partial F}{\partial\theta_k}\right|_{\boldsymbol{\theta}^0} + \left(y_i - F\left(\boldsymbol{\theta}^0, x_i\right)\right)\left.\frac{\partial^2 F}{\partial\theta_j \partial\theta_k}\right|_{\boldsymbol{\theta}^0}\right). \qquad (4)$$

Обозначим градиент целевой функции как $\mathbf{g}$ и матрицу Гессе как $\mathbf{H}$, которые определяются как

$$\mathbf{g} \equiv \left[g_i\right] = \left[\left.\frac{\partial E}{\partial\theta_j}\right|_{\boldsymbol{\theta}^0}\right], \qquad (5)$$

$$\mathbf{H} \equiv \left[h_{jk}\right] = \left[\left.\frac{\partial^2 E}{\partial\theta_j \partial\theta_k}\right|_{\boldsymbol{\theta}^0}\right]. \qquad (6)$$

В результате уравнение (2) может быть записано в матричной форме

$$E^* = E\big|_{\boldsymbol{\theta}^0} + \left(\delta\boldsymbol{\theta}\right)^{\mathrm{T}}\mathbf{g} + \frac{1}{2}\left(\delta\boldsymbol{\theta}\right)^{\mathrm{T}}\mathbf{H}\left(\delta\boldsymbol{\theta}\right), \qquad (7)$$

где $\delta\mathbf{и} = \mathbf{и} - \mathbf{и}^0$.

Здесь E^* – аппроксимация по методу Ньютона значения E в окрестности $\mathbf{и}^0$. Минимизация E^* требует, чтобы ее производная относительно $\delta\mathbf{и}$ равнялась нулю в неподвижной точке:

$$\frac{\partial E^*}{\partial\left(\delta\mathbf{и}\right)} = 0. \qquad (8)$$

Подстановка (7) в (8) приводит к следующему:

$$\delta\mathbf{и} = -\mathbf{H}^{-1}\mathbf{g}. \qquad (9)$$

$\delta\mathbf{и}$, полученное из (9), является просто условием экстремума для E^*. Поэтому требуются итерации для нахождения решения для E.

Новое решение определяется по формуле

$$\mathbf{и} = \mathbf{и}^0 + \delta\mathbf{и}. \qquad (10)$$

Итеративная процедура повторяется, пока не будет достигнута желаемая сходимость. Необходимо учесть, что когда $\delta\mathbf{и}$ станет пренебрежимо мало, то значение E^* в неподвижной точке практически совпадает с E, то есть целевая функция достигает экстремума. С методом Ньютона связаны две очень важные проблемы, которые возникают в зависимости от природы функции, используемой для описания модели. Первая трудность касается членов, содержащих вторую производную, которые расположены на главной диагонали матрицы Гессе. В результате нет гарантии, что матрица будет положительно определенной, а значит, совсем необязательно, что на каждом шаге итерации новое решение будет стремиться к точке минимума. Вторая проблема связана с тем, что итерационный процесс может сходиться очень медленно или даже расходиться в случае, если матрица Гессе плохо обусловлена. Причинами могут быть сильная корреляция, наблюдаемая между некоторыми параметрами, и нечувствительность функции, описывающей модель, к определенным параметрам. Поэтому необходим ряд модификаций метода для преодоления этих трудностей, чтобы обеспечить быструю сходимость к точке минимума. В методе Гаусса члены, содержащие вторую производную, отбрасываются, как если бы они были равны нулю. Обычно такая аппроксимация является удовлетворительной из-за нулевого градиента целевой функции в точке минимума. Такая модификация делает матрицу Гессе положительно

определенной и гарантирует сходимость к точке минимума. Метод Маквардта полезен, когда матрица Гессе в методе Гаусса плохо обусловлена. Добавление константы к диагональным элементам этой матрицы улучшают ее обусловленность и препятствуют численной вырожденности матрицы Гессе. Специальная техника интерполяции и экстраполяции изменяют длину шага для повышения скорости сходимости. Этот способ известен также как линейный поиск. Штрафные функции могут использоваться для дополнительного улучшения скорости сходимости путем ограничения области поиска.

В итоге значение целевой функции в точке минимума вычисляется следующим образом:

$$E\big|_{\text{и}} = E\big|_{\hat{\text{и}}} + \sum_{j=1}^{m}\sum_{k=1}^{m}\left(\theta_j - \hat{\theta}_j\right)\left(\theta_k - \hat{\theta}_k\right)\frac{\partial^2 E}{\partial\theta_j \partial\theta_k}\bigg|_{\hat{\text{и}}}. \tag{11}$$

Это выражение эквивалентно следующей записи:

$$F(\text{и}) = F\left(\hat{\text{и}}\right) + \sum_{j=1}^{m}\left(\frac{\partial F}{\partial\theta_j}\right)_{\text{и}=\hat{\text{и}}} \cdot \left(\theta_j - \hat{\theta}_j\right). \tag{12}$$

Статистический вывод

Со статистической точки зрения неявно предполагается, что функция, описывающая модель, имеет следующую форму:

$$y_i = F(\text{и}, x_i) + \varepsilon_i, \tag{13}$$

где ε_i – случайная ошибка, характеризующаяся своей функцией плотности распределения вероятностей.

Оценка по МНК широко используется для решения обратных задач, т.к. по центральной предельной теореме сумма независимых случайных величин с конечной дисперсией стремится также быть распределенной по нормальному закону. Для нелинейной модели вышесказанное уже, в общем-то, неверно. Однако считается, что для области в пространстве параметров достаточно близкой к оценкам нелинейную функцию, описывающую модель, можно аппроксимировать линейной функцией путем разложения в ряд Тейлора первого порядка. Это допущение остается справедливым для большинства моделей пласта до тех пор, пока используется большой объем данных. В итоге для больших n даже в случае нелинейной регрессии оценка по МНК $\hat{\text{и}}$ может выступать в качестве несмещенной оценки и с минимальной дисперсией.

Доверительные интервалы

Непосредственное применение доверительных интервалов к результатам гидродинамических исследований скважин (ГДИС) требует выполнения двух условий. Во-первых, ошибки, которые представляют собой разность между действительным значением давления и его истинной величиной, должны быть независимы и нормально распределены относительно истинного изменения давления. Это предположение образует статистический базис МНК. Второе условие состоит в том, чтобы для области в пространстве параметров достаточно близкой к их оценкам целевую функцию можно было аппроксимировать линейной формой путем разложения в ряд Тейлора первого порядка.

Условие возможности аппроксимации функции, описывающей модель, путем разложения ее в ряд Тейлора первого порядка приводит к следующему выражению:

$$F(\mathbf{и}) = F(\hat{\mathbf{и}}) + \sum_{j=1}^{m} \left(\frac{\partial F}{\partial \theta_j} \right)_{\mathbf{и}=\hat{\mathbf{и}}} \left(\theta_j - \hat{\theta}_j \right). \tag{14}$$

Считается, что наблюдаемые данные замеров давления y_i нормально распределены относительно истинного значения $F(\mathbf{и},\ x_i)$ с известной дисперсией:

$$Prob\left(y_i \middle| F(\mathbf{и},\ x_i) \right) = Prob\left(y_i \middle| \mathbf{и} \right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} \left(y_i - F(\mathbf{и},\ x_i) \right)^2 \right). \tag{15}$$

В результате наблюдений n значений давления функция правдоподобия для параметров имеет вид

$$\begin{aligned} L\left(\mathbf{и} \middle| y_1, \ldots, y_n \right) &= Prob\left(y_1, \ldots, y_n \middle| \mathbf{и} \right) = Prob\left(y_1 \middle| \mathbf{и} \right) \cdot \ldots \cdot Prob\left(y_n \middle| \mathbf{и} \right) = \\ &= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} \left(y_i - F(\mathbf{и},\ x_i) \right)^2 \right) \right) = \\ &= \frac{1}{\left(\sqrt{2\pi}\sigma \right)^n} \exp\left(-\frac{1}{2\sigma^2} \left(\mathbf{R} - \mathbf{J} \cdot \left(\mathbf{и} - \hat{\mathbf{и}} \right) \right)^{\mathrm{T}} \left(\mathbf{R} - \mathbf{J} \cdot \left(\mathbf{и} - \hat{\mathbf{и}} \right) \right) \right), \end{aligned} \tag{16}$$

где

$$\mathbf{R} = \begin{pmatrix} y_1 - F(\hat{\mathbf{и}},\ x_1) \\ \vdots \\ y_n - F(\hat{\mathbf{и}},\ x_n) \end{pmatrix} \tag{17}$$

и

$$\mathbf{J} = \begin{pmatrix} \left(\frac{\partial F}{\partial \theta_1} \right)_{x_1} & \cdots & \left(\frac{\partial F}{\partial \theta_m} \right)_{x_1} \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial F}{\partial \theta_1} \right)_{x_n} & \cdots & \left(\frac{\partial F}{\partial \theta_m} \right)_{x_n} \end{pmatrix}_{\mathbf{и}=\hat{\mathbf{и}}}. \tag{18}$$

МНК эквивалентен максимуму функции правдоподобия, который имеет место тогда и только тогда, когда

$$\mathbf{R}^{\mathrm{T}} \mathbf{J} = \mathbf{0}. \tag{19}$$

В результате

$$L\left(\mathbf{и} \middle| y_1, \ldots, y_n \right) = \frac{1}{\left(\sqrt{2\pi}\sigma \right)^n} \exp\left(-\frac{1}{2\sigma^2} \left(\mathbf{R}^{\mathrm{T}} \mathbf{R} + \left(\mathbf{и} - \hat{\mathbf{и}} \right)^{\mathrm{T}} \mathbf{J}^{\mathrm{T}} \mathbf{J} \left(\mathbf{и} - \hat{\mathbf{и}} \right) \right) \right). \tag{20}$$

Матрица Гессе в методе Гаусса, поделенная на 2, определяется как

$$\mathbf{H}=\left.\begin{pmatrix} \sum\limits_{i=1}^{n}\left(\frac{\partial F}{\partial \theta_1}\right)\left(\frac{\partial F}{\partial \theta_1}\right) & \cdots & \sum\limits_{i=1}^{n}\left(\frac{\partial F}{\partial \theta_1}\right)\left(\frac{\partial F}{\partial \theta_m}\right) \\ \vdots & \ddots & \vdots \\ \sum\limits_{i=1}^{n}\left(\frac{\partial F}{\partial \theta_m}\right)\left(\frac{\partial F}{\partial \theta_1}\right) & \cdots & \sum\limits_{i=1}^{n}\left(\frac{\partial F}{\partial \theta_m}\right)\left(\frac{\partial F}{\partial \theta_m}\right) \end{pmatrix}\right|_{\mathbf{и}=\hat{\mathbf{и}}} . \tag{21}$$

Тогда

$$\mathbf{J}^{\mathrm{T}}\mathbf{J}=\mathbf{H}. \tag{22}$$

Если для параметров используется локально равномерное априорное распределение вероятностей (неинформативное априорное распределение вероятностей), тогда по теореме Байеса апостериорное распределение вероятностей параметров после n наблюдений:

$$Prob\left(\mathbf{и}\middle| y_1,\ldots,y_n\right)=\frac{L\left(\mathbf{и}\middle| y_1,\ldots,y_n\right)\cdot Prob\left(\mathbf{и}\right)}{\int\limits_{m} L\left(\mathbf{и}\middle| y_1,\ldots,y_n\right)\cdot Prob\left(\mathbf{и}\right)d\mathbf{и}}=\frac{L\left(\mathbf{и}\middle| y_1,\ldots,y_n\right)}{\int\limits_{m} L\left(\mathbf{и}\middle| y_1,\ldots,y_n\right)d\mathbf{и}}, \tag{23}$$

где $Prob\left(\mathbf{и}\right)$ – локально равномерное априорное распределение вероятностей.

По определению многомерного нормального распределения:

$$\int\limits_{m}\frac{\left|\mathbf{H}\right|^{1/2}}{\left(2\pi\sigma^2\right)^{m/2}}\exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{и}-\hat{\mathbf{и}}\right)^{\mathrm{T}}\mathbf{H}\left(\mathbf{и}-\hat{\mathbf{и}}\right)\right)d\mathbf{и}=1. \tag{24}$$

Поэтому уравнение принимает вид

$$Prob\left(\mathbf{и}\middle| y_1,\ldots,y_n\right)=\frac{\left|\mathbf{H}\right|^{1/2}}{\left(2\pi\sigma^2\right)^{m/2}}\exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{и}-\hat{\mathbf{и}}\right)^{\mathrm{T}}\mathbf{H}\left(\mathbf{и}-\hat{\mathbf{и}}\right)\right). \tag{25}$$

То есть параметры $\mathbf{и}$ образуют многомерное нормальное распределение относительно $\hat{\mathbf{и}}$ с матрицей ковариаций $\sigma^2\mathbf{H}^{-1}$. Уравнение (25) количественно выражает неопределенность, связанную с оценками параметров.

Когда дисперсия σ^2 неизвестна, вышеописанные рассуждения требуют небольшого уточнения. σ^2 можно получить на основе среднего квадрата ошибок s^2, который вычисляется как

$$s^2=\frac{SS_{\mathrm{err}}}{n-m}, \tag{26}$$

где

$$SS_{\mathrm{err}}=\sum_{i=1}^{n}\left(y_i-F\left(\mathbf{и},\, x_i\right)\right)^2. \tag{27}$$

В этом случае s^2 – это несмещенная оценка σ^2, и σ^2 имеет обратное гамма-распределение относительно s^2 с $n-m$ степенями свободы:

$$Prob\left(\sigma^2\middle| s^2\right)=\frac{2}{\Gamma\left(\nu/2\right)}\left(\frac{\nu s^2}{2}\right)\frac{1}{\sigma^{\nu+1}}\exp\left(-\frac{\nu s^2}{2\sigma^2}\right), \tag{28}$$

где $\nu=n-m$.

Поскольку $\mathbf{и}$ и σ^2 – независимые случайные величины, то $\hat{\mathbf{и}}$ не меняется, даже когда σ^2 заменяется s^2.

В итоге апостериорное распределение вероятностей для **и** может быть получено путем исключения σ^2 при интегрировании суммарной апостериорной плотности распределения вероятностей для **и** и σ^2:

$$Prob\left(u \middle| y_1,\ldots,y_n\right)=\int_0^{\infty} Prob\left(u,\ \sigma^2 \middle| y_1,\ldots,y_n\right)d\sigma^2=$$
$$=\int_0^{\infty} Prob\left(u \middle| y_1,\ldots,y_n,\ \sigma^2\right)\cdot Prob\left(\sigma^2 \middle| y_1,\ldots,y_n\right)d\sigma^2= \qquad (29)$$
$$=\int_0^{\infty} Prob\left(uy_1,\ldots,y_n,\ \sigma^2\right)\cdot Prob\left(\sigma^2 \middle| s^2\right)d\sigma^2.$$

После подстановки (25) и (28) в (29) получим

$$Prob\left(\mathbf{и} \middle| y_1,\ldots,y_n\right)=\frac{\Gamma(n/2)\left|\mathbf{H}\right|^{1/2}s^{-m}}{\left(\Gamma(1/2)\right)^m\Gamma(v/2)\left(\sqrt{v}\right)^m}\left(1+\frac{\left(\mathbf{и}-\hat{\mathbf{и}}\right)^{\mathrm{T}}\mathbf{H}\left(\mathbf{и}-\hat{\mathbf{и}}\right)}{vs^2}\right)^{-\frac{v+m}{2}}. \qquad (30)$$

Поэтому, когда σ^2 неизвестна, параметры **и** образуют многомерное t-распределение Стьюдента относительно $\hat{\mathbf{и}}$ с матрицей ковариаций $s^2\mathbf{H}^{-1}$ и $n-m$ степенями свободы.

Маргинальное распределение вероятностей параметра θ_j определяется путем исключения θ_i ($i\neq j$, $i=1,\ldots,m$) при интегрировании по пространству **и**:

$$Prob\left(\theta_j \middle| y_1,\ldots,y_n\right)=\frac{1}{\sqrt{2\pi\sigma_{\theta_j}^2}}\exp\left(-\frac{1}{2\sigma_{\theta_j}^2}\left(\theta_j-\hat{\theta}_j\right)^2\right), \qquad (31)$$

где σ_{θ_j} – СКО, определяемое как

$$\sigma_{\theta_j}^2=\sigma^2h_{jj}^{-1}, \qquad (32)$$

где h_{jj}^{-1} – j-й диагональный элемент обратной матрицы Гессе, вычисляемой в точке $\mathbf{и}=\hat{\mathbf{и}}$.

По определению, 95 %-й доверительный интервал накрывает 95 % площади под кривой плотности распределения вероятностей, т.е. он представляет собой диапазон, доверительная вероятность попадания значений параметра внутрь которого составляет 95 %. Обычно используются два типа доверительных интервалов: диапазон абсолютных значений и диапазон относительных значений. Относительные значения получаются путем деления абсолютных значений на величину оценки параметра.

В случаях, когда дисперсия σ^2 неизвестна, $(1-\alpha)\cdot 100$ %-й доверительный интервал для каждого параметра определяется из следующего неравенства:

$$\hat{\theta}_j-\sigma_{\theta_j}\cdot t_{1-\alpha/2}\leq\theta_j\leq\hat{\theta}_j+\sigma_{\theta_j}\cdot t_{1-\alpha/2}, \qquad (33)$$

где $t_{1-\alpha/2}$ – табличное значение квантиля порядка $1-\alpha/2$ для t-распределения Стьюдента с $n-m$ степенями свободы.

В случаях, когда $n-m>30$, значение $t_{1-\alpha/2}$ может быть заменено соответствующим значением для нормального распределения. Так, для $\alpha=0,05$ его значение будет равно $1,96$. Тогда (33) примет вид

$$\hat{\theta}_j-1,96\cdot\sigma_{\theta_j}\le\theta_j\le\hat{\theta}_j+1,96\cdot\sigma_{\theta_j}. \quad (34)$$

$(1-\alpha)\cdot 100$ %-й доверительный интервал для относительных значений каждого параметра определяется из следующего неравенства:

$$1-\frac{\sigma_{\theta_j}\cdot t_{1-\alpha/2}}{\hat{\theta}_j}\le\frac{\theta_j}{\hat{\theta}_j}\le 1+\frac{\sigma_{\theta_j}\cdot t_{1-\alpha/2}}{\hat{\theta}_j}. \quad (35)$$

Аналогично, когда $n-m>30$, доверительный интервал для относительных значений каждого параметра можно представить в виде

$$1-\frac{1,96\cdot\sigma_{\theta_j}}{\hat{\theta}_j}\le\frac{\theta_j}{\hat{\theta}_j}\le 1+\frac{1,96\cdot\sigma_{\theta_j}}{\hat{\theta}_j}. \quad (36)$$

На практике используются значения доверительных интервалов, приведенные в табл. 1 [1]. Они были получены эвристическим путем на основе реальных экспериментов по интерпретации промысловых и смоделированных данных ГДИС. Ключевой идеей является то, что если модель выбрана правильно и присутствует достаточное количество данных, то все параметры должны быть внутри этих допустимых пределов.

Таблица 1 – Приемлемые доверительные интервалы по [1]

Параметры	Диапазон относительных значений	Диапазон абсолютных значений
Проницаемость (k)	±10 %	
Коэффициент накопления (C)	±10 %	
Расстояние до границы (r_e)	±10 %	
Длина трещины (x_f)	±10 %	
Коэффициент упругоемкости (ω)	±20 %	
Коэффициент пропускания (λ)	±20 %	
Скин-фактор (S)		±1
Первоначальное давление (P_i)		±0,005 МПа

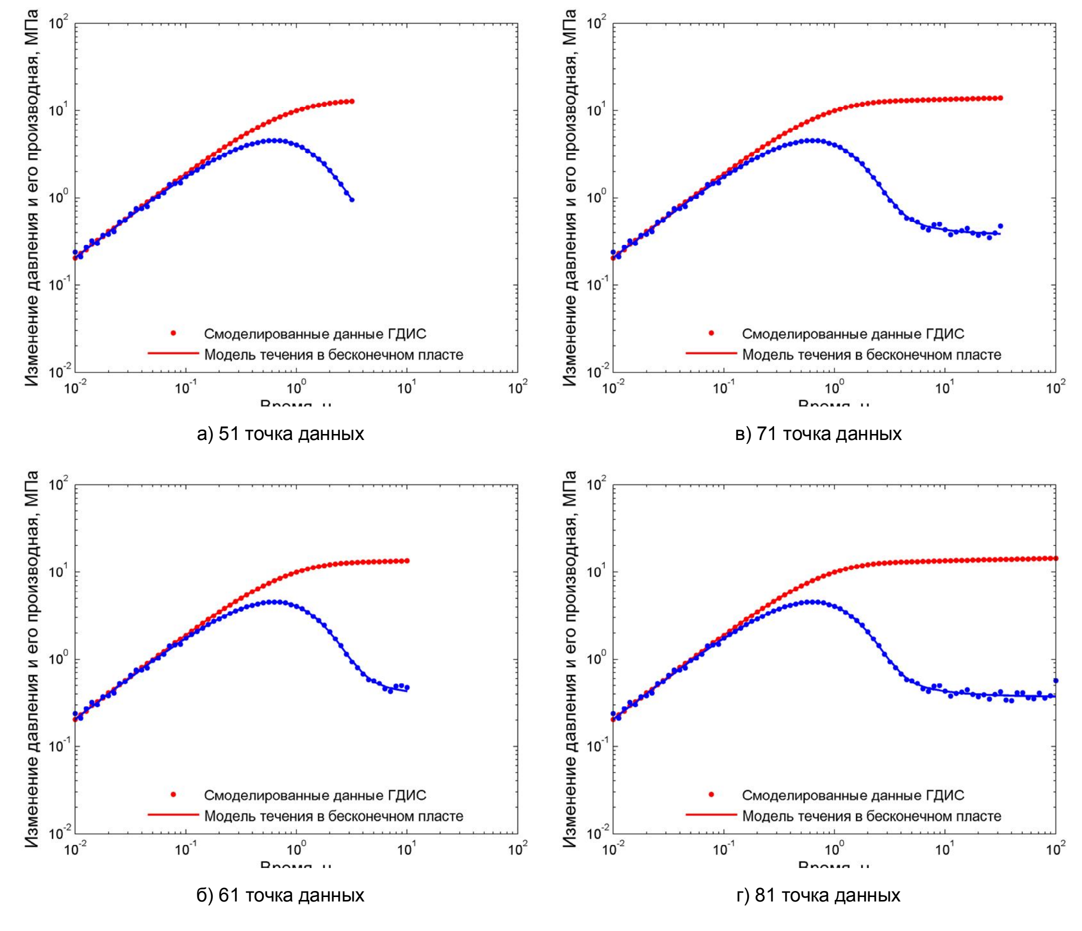
Продемонстрируем, как можно использовать доверительные интервалы для оценки корректности модели. Для этого были смоделированы данные ГДИС методом понижения уровня. Цель демонстрации – показать, как доверительные интервалы решают задачу, когда заранее известно, соответствует ли модель пласта данным или нет.

В первом случае модель была выбрана правильно. Значения давления при ГДИС методом понижения уровня рассчитывались с использованием модели течения в бесконечном пласте, к которым затем прибавлялись случайные ошибки. Информация о пласте и насыщающем его флюиде представлена в табл. 2. Истинные значения параметров $k=0,05$ мкм^2, $S=10$ и $C=0,2$ м^3/МПа. Генератор случайных чисел

формировал набор случайных ошибок, распределенных по нормальному закону с нулевым математическим ожиданием и дисперсией $2{,}5\cdot 10^{-5}$ МПа^2. В зависимости от числа точек данных рассматривались следующие четыре случая: а) 51 точка данных, б) 61 точка данных, в) 71 точка данных и г) 81 точка данных. Применялась модель течения в бесконечном пласте с тремя параметрами (k, S и C). Соответствие модели данным иллюстрируется на рис. 1.

Таблица 2 – Информация о пласте и насыщающем его флюиде

Радиус ствола скважины (r_w)	м	0,1
Мощность пласта (h)	м	5
Объемный коэффициент (B_o)	$\left[м^3\right]_{пласт}/\left[м^3\right]_{норм}$	1
Вязкость (μ)	Па·с	10^{-3}
Пористость (ϕ)		0,2
Первоначальное давление (P_i)	МПа	20
Общая сжимаемость (c_t)	МПа$^{-1}$	10^{-4}
Рабочий дебит (q)	м^3/сут	100



а) 51 точка данных

в) 71 точка данных

б) 61 точка данных

г) 81 точка данных

Рисунок 1 – Смоделированные данные ГДИС
и соответствие им корректно выбранной модели пласта

Таблица 3 – 95 %-е доверительные интервалы для проницаемости в случае корректной модели

	а)	б)	в)	г)
Число точек данных	51	61	71	81
Оценка параметра	0,0484	0,0498	0,0498	0,0499
s^2	$2,74\cdot10^{-5}$	$2,90\cdot10^{-5}$	$2,67\cdot10^{-5}$	$2,69\cdot10^{-5}$
h^{-1}	$1,55\cdot10^{-2}$	$2,09\cdot10^{-3}$	$6,40\cdot10^{-4}$	$2,82\cdot10^{-4}$
$\sigma_\theta^2 = s^2h^{-1}$	$4,24\cdot10^{-7}$	$6,07\cdot10^{-8}$	$1,71\cdot10^{-8}$	$7,57\cdot10^{-9}$
σ_θ	$6,51\cdot10^{-4}$	$2,46\cdot10^{-4}$	$1,31\cdot10^{-4}$	$8,70\cdot10^{-5}$
Доверительный интервал	2,71 %	0,99 %	0,52 %	0,35 %
Решение	приемлемо	приемлемо	приемлемо	приемлемо

Для простоты приводятся результаты по оценкам только для одного параметра – проницаемости. Маргинальные распределения вероятностей изображены на рис. 2. Соответствующие 95 %-е доверительные интервалы для проницаемости сведены в табл. 3.

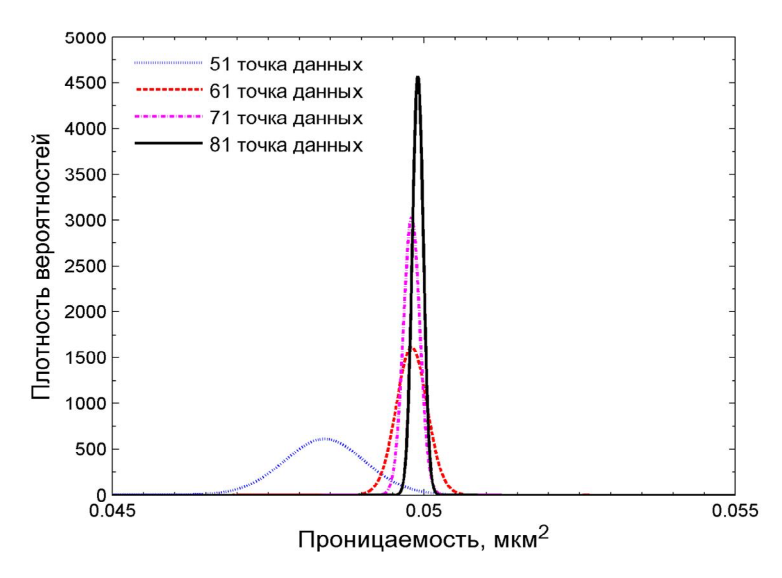


Рисунок 2 – Маргинальные плотности распределения вероятностей в случае корректно выбранной модели пласта

По сути, только случаи б), в) и г) содержат полезную информацию о проницаемости. Как следует из табл. 3, оценки проницаемости достаточно близки к истинному значению ($k = 0,05$ мкм^2). Поэтому на рис. 2 все распределения вероятностей группируются вокруг этого значения. С увеличением числа данных появляется больше информации о проницаемости, и соответствующее ей СКО (σ_θ) уменьшается. Размах распределений сужается, и нормальное распределение стремится принять форму дельта-функции Дирака. С позиций доверительных интервалов все случаи являются приемлемыми, т.е. модель выбрана верно.

Следующий случай соответствует некорректной с точки зрения данных модели. Данные ГДИС методом понижения уровня были получены с помощью модели непроницаемой внешней границы, на которые затем накладывались такие же, как и в предыдущем случае, случайные ошибки. Истинные значения параметров $k = 0,05$ мкм^2, $S = 10$, $C = 0,2$ м^3/МПа и $r_e = 500$ м. Для подгонки кривой изменения давления применялась модель течения в бесконечном пласте с тремя параметрами (k, S и C). Соответствие модели данным иллюстрируется на рис. 3.

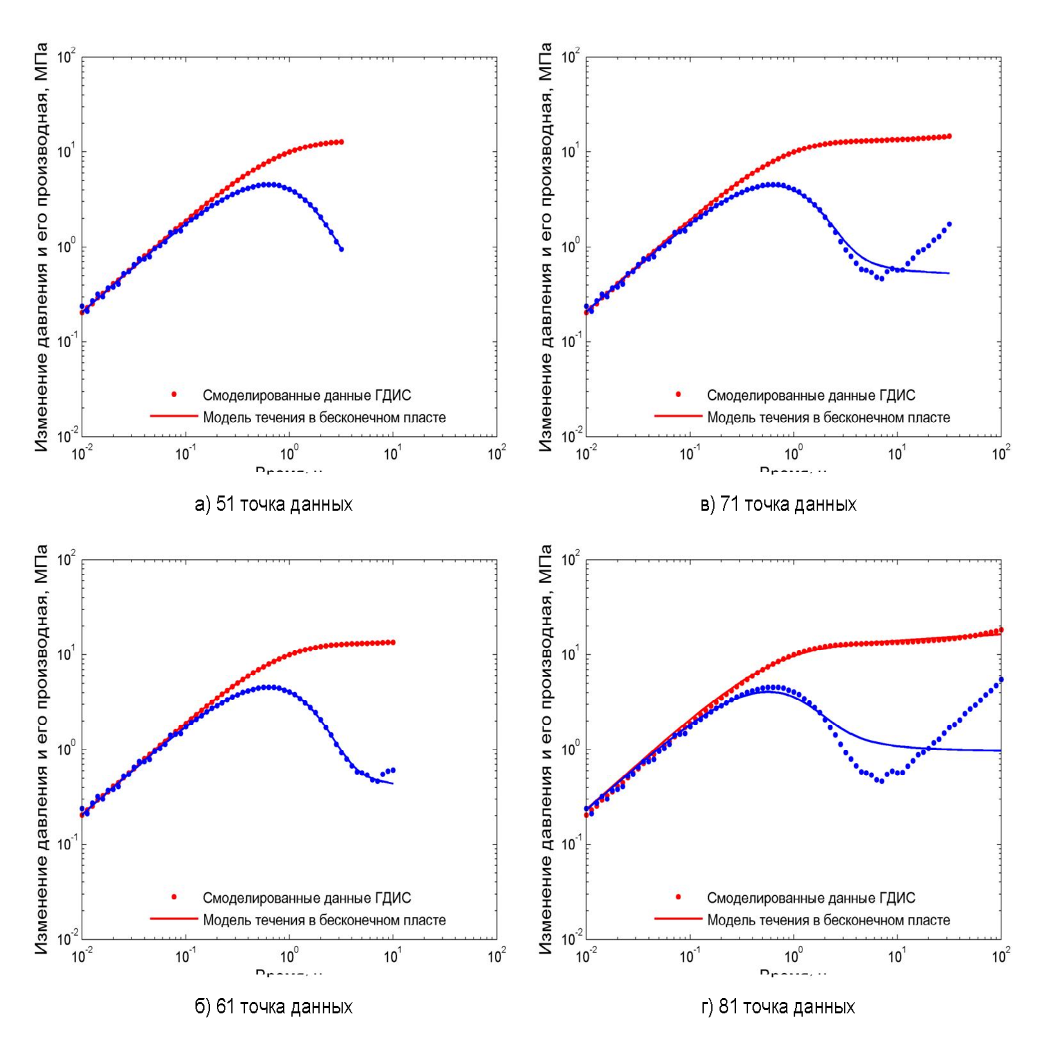


Рисунок 3 – Смоделированные данные ГДИС и соответствие им некорректно выбранной модели пласта

Маргинальные распределения вероятностей изображены на рис. 4. Соответствующие 95 %-е доверительные интервалы для проницаемости сведены в табл. 4.

Таблица 4 – 95 %-е доверительные интервалы для проницаемости в случае некорректной модели

	а)	б)	в)	г)
Число точек данных	51	61	71	81
Оценка параметра	0,0484	0,0487	0,0362	0,019
s^2	$2,74\cdot10^{-5}$	$4,77\cdot10^{-5}$	$6,59\cdot10^{-3}$	$1,98\cdot10^{-1}$
h^{-1}	$1,55\cdot10^{-2}$	$1,92\cdot10^{-3}$	$1,80\cdot10^{-4}$	$6,10\cdot10^{-6}$
$\sigma_\theta^2 = s^2h^{-1}$	$4,24\cdot10^{-7}$	$9,14\cdot10^{-8}$	$1,18\cdot10^{-6}$	$1,21\cdot10^{-6}$
σ_θ	$6,51\cdot10^{-4}$	$3,02\cdot10^{-4}$	$1,09\cdot10^{-3}$	$1,10\cdot10^{-3}$
Доверительный интервал	2,71 %	1,24 %	6,00 %	11,49 %
Решение	приемлемо	приемлемо	приемлемо	неприемлемо

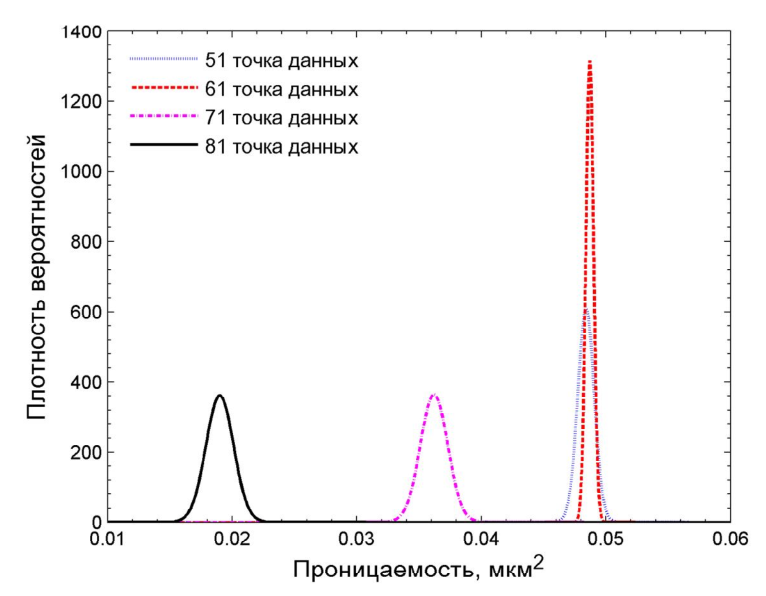


Рисунок 4 – Маргинальные плотности распределения вероятностей в случае некорректно выбранной модели пласта

В случае г) и д) модель течения в бесконечном пласте не соответствует данным. По сравнению со случаями а) и б) средний квадрат ошибок s^2 для случаев в) и г) начинает возрастать, что приводит к увеличению размаха графиков плотностей распределения вероятностей. Если случаи а), б) и в) еще приемлемы с позиций доверительных интервалов, то случай г) уже неприемлем, что говорит о некорректном выборе модели. В принципе доверительные интервалы можно использовать для принятия или отклонения выбранной модели. Независимо от того, корректно или нет выбрана модель, доверительные интервалы в конечном счете дают непротиворечивые результаты. Но нужно учитывать, что на практике при верификации модели доверительные интервалы следует определять для всех параметров. Кроме того, доверительные интервалы легко вычислять, поскольку вся необходимая информация содержится в результатах нелинейной регрессии, и нетрудно использовать для верификации модели, что было продемонстрировано выше. Однако сравнительный анализ на основе доверительных интервалов имеет два недостатка (практический и теоретический) с точки зрения дискриминантного анализа моделей.

Во-первых, доверительные интервалы прямо пропорциональны дисперсии распределения вероятностей параметра, который в свою очередь представляет собой комбинацию среднего квадрата ошибок (оценочная дисперсия) s^2 и диагонального элемента обратной матрицы Гессе h_{jj}^{-1}. Например, в случае г) (табл. 4) оценка дисперсии составляет $6{,}59 \cdot 10^{-3}$ МПа^2, что намного больше истинной дисперсии $2{,}5 \cdot 10^{-5}$ МПа^2. Однако из-за того, что диагональный элемент обратной матрицы Гессе относительно мал ($1{,}8 \cdot 10^{-4}$) вследствие большого числа точек данных (71 точка данных), соответствующая дисперсия для проницаемости составила всего $1{,}18 \cdot 10^{-6}$. В результате, 95 %-й доверительный интервал оказался приемлемым (6 %). Тем не менее, анализируя рис. 3, можно предположить о существовании непроницаемой границы, т.к. нетрудно различить крутой наклон прямой на графике производной функции изменения давления. То есть доверительные интервалы могут находиться в допустимых пределах, даже если используется некорректная модель.

Во-вторых, доверительные интервалы удобны при верификации моделей, но не подходят для их дискриминантного анализа. Другими словами, на основе доверительных интервалов можно определить, подходит ли модель или нет, но ничего нельзя сказать о том, какая из моделей лучше. Это связано с тем, что при вычислении доверительных интервалов не учитывается корреляция между параметрами. Однако в общем случае параметры пласта нелинейно связаны друг с другом, что необходимо принимать в расчет при верификации. Более того, уравнение (25) говорит о том, что размерность распределения вероятностей параметров совпадает с их количеством. То есть различные модели с различным числом параметров имеют различные размерности распределений вероятностей. Поэтому прямое сравнение соответствующих доверительных интервалов явно недостаточно.

Литература

1. Anraku T. Discrimination between reservoir models in well test analysis, Ph.D. Thesis. – Stanford University, 1993.
2. Денисов С.В. Оценка параметров и идентификация модели пласта по результатам гидродинамических исследований скважин // Вестник Московской Академии рынка труда и информационных технологий. – 2005. – № 9 (21). – С. 106-119.
3. Horne R.N. Modern well test analysis: a computer-aided approach. – 4th printing, Petroway. – 1990.

С.В. Денисов, К.А. Сидельников

Верифікація моделі пласта-колектора за даними гідродинамічних досліджень свердловин з використанням довірчих інтервалів

Представлена базова процедура нелінійного регресійного аналізу. Використаний байєсовський висновок, при якому вся інформація про параметри пласта виражається у термінах ймовірності, а невизначеність, пов’язана з оцінками цих параметрів, може визначатися кількісним чином. Досліджені проблеми, які виникають під час використання довірчих інтервалів у межах байєсовського висновку для дискримінантного аналізу різних моделей-кандидатів.

S.V. Denisov, K.A. Sidelnikov

The basic procedures of nonlinear regression are presented. Bayesian inference is introduced. In Bayesian inference all information about the reservoir parameters is expressed in terms of probability, and uncertainty involved in the parameter estimates can be expressed quantitatively. Confidence intervals are derived in the framework of Bayesian inference, and the problems inherently involved in the application of confidence intervals for model discrimination are discussed.

Статья поступила в редакцию 02.07.2008.